OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.3

Run on: January 24, 2003, 11.42.22

Title: us-09-910-428-2
Perfect score: 26
Sequence: 1 ctctcccaaaltcgilargttrt 26

Scoring table: IDENTITY_NUC
                Caper 10.0 , caper 1.0

Searched: 205,640 seqs 1455140,974 residues

Total number of hits satisfying chosen paramters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing : Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

, Search time 66.3129 Seconds
(without alignments)
11410.644 Million cell updates/sec

Database :   GenEmbl:*
    1: gb_ba:*
    2: gb_htg:*
    3: gb_in:*
    4: gb_om:*
    5: gb_cw:*
    6: gb_rat:*
    7: gb_ph:*
    8: gb_pl:*
    9: gb_pr:*
   10: gb_ro:*
   11: gb_sts:*
   12: gb_sy:*
   13: gb_vi:*
   14: qb_vi:*
   15: em_ba:*
   16: em_fun:*
   17: em_hum:*
   18: em_in:*
   19: em_mu:*
   20: em_om:*
   21: em_or:*
   22: em_ov:*
   23: cm_pl:*
   24: em_ph:*
   25: em_ro:*
   26: em_sts:*
   27: em_un:*
   28: em_vi:*
   29: em_vrl:*
   30: em_htg_hum:*
   31: em_htg_inv:*
   32: em_htg_other:*
   33: em_htg_mus:*
   34: em_htg_plu:*
   35: em_htg_rod:*
   36: em_htg_mam:*
   37: em_htg_vrt:*
   38: em_sy:*
   39: em_htgo_hum:*
   40: em_htgo_mus:*
   41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		26	100.0	732	4	AF040955	AF040955 Bos indic
2		26	100.0	26588	4	H1015731	U15731 Bos tauru
3		22.8	87.7	349	4	AF126288	AF126288 Bos tauru
4	c	21.4	82.3	14569	2	AC125465	AC125465 Mus muscu
5	c	21.4	82.3	94161	2	AC020937	AC020937 Homo sapi
6	c	21.4	82.3	123495	2	AC008502	AC008502 Homo sapi
7	c	21.4	82.3	164478	2	AC008787	AC008787 Homo sapi
8	c	21.4	82.3	195217	2	AC124754	AC124754 Mus muscu
9	c	21.4	81.5	186552	2	AC069033	AC069033 Rattus no
10		21.2	80.8	199287	8	AC123369	AC123369 Homo sapi
11		20.4	77.7	110600	8	AF011078	AF011078 Homo sapi
12		20.4	77.7	157563	2	AF004470	AF004470 Lotus jap
13	c	20.2	77.2	230419	2	AC073450	AC073450 Homo sapi
14	c	20.2	77.2	27804	5	AC107804	AC107804 Mus muscu
15		19.8	76.2	95445	2	AC016175	AC016175 Homo sapi
16	c	19.8	76.2	99601	2	AL354622	AL354622 Homo sapi
17		19.8	76.2	99888	2	AC025518	AC025518 Homo sapi
18	c	19.8	76.2	124839	9	AC097109	AC097109 Homo sapi
19	c	19.8	76.2	127917	9	HS257A7	HS257A7 Human DNA
20		19.8	76.2	147945	2	AC084735	AC084735 Homo sapi
21		19.8	76.2	163696	9	AC108616	AC108616 Rattus no
22		19.8	76.2	170776	9	AL139087	AL139087 Human DNA
23		19.8	76.2	176129	9	AC124533	AC124533 Mus muscu
24	c	19.8	76.2	183530	9	AL160553	AL160553 Mus muscu
25		19.8	76.2	188488	2	AC125197	AC125197 Human DNA
26	c	19.8	76.2	195278	2	AL139779	AL139779 Homo sapi
27	c	19.8	76.2	196145	2	AC126241	AC126241 Homo sapi
28	c	19.8	76.2	202991	2	AC125178	AC125178 Mus muscu
29	c	19.8	76.2	204754	2	AC125068	AC125068 Mus muscu
30	c	19.8	76.2	211342	2	AC124589	AC124589 Mus muscu
31		19.8	76.2	217320	2	AC121575	AC121575 Mus muscu
32		19.8	76.2	249608	2	AC124416	AC124416 Mus muscu
33	c	19.4	75.2	25387	2	AC140832	AC140832 Mus muscu
34	c	19.4	75.2	104424	2	AC095918	AC095918 Rattus no
35	c	19.6	75.4	115043	2	AL663070	AL663070 Human DNA
36	c	19.6	75.4	121156	9	AC127075	AC127075 Rattus no
37		19.6	75.4	136159	9	AC097475	AC097475 Homo sapi
38		19.6	75.4	136159	9	AC068636	AC068636 Homo sapi
39		19.6	75.4	136730	9	AC108110	AC108110 Homo sapi
40		19.6	75.4	154399	2	AC127822	AC127822 Rattus no
41		19.6	75.4	161408	2	AC097910	AC097910 Rattus no
42	c	19.6	75.4	163549	2	AC121648	AC121648 Rattus no
43		19.6	75.4	191191	2	AC115524	AC115524 Rattus no
44		19.6	75.4	195430	2	AC263347	AC263347 Homo sapi
45		19.6	75.4	200484	2	AC025002	AC025002 Homo sapi

ALIGNMENTS

RESULT 1
AF040955/c
LOCUS   AF040955   732 bp   DNA   linear   MAM 16-JAN-1998
DEFINITION Bos indicus growth hormone receptor promoter and exon 1, partial
            sequence.
ACCESSION   AF040955
VERSION   AF040955.1 GI:2773337
KEYWORDS   .
SOURCE   Bos indicus
  ORGANISM Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 732)
  AUTHORS   Lucy,M.C. and Boyd,C.K.

TITLE   Promoter and first exon for bovine growth hormone receptor isolated
            from Bos indicus
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 742)
AUTHORS   Lucy,M.C. and Boyd,C.K.
TITLE   Direct Submission
JOURNAL   Submitted (02-JAN-1998) Animal Science, University of Missouri, 164
            ASRC, Columbia, MO 65211, USA
FEATURES   Location/Qualifiers
     source   1..742
                     /organism="Bos indicus"
                     /strain="Brahman"
                     /db_xref="taxon:9915"
                     /chromosome="20"
     promoter   1..569
                     /note="growth hormone receptor promoter"
     mRNA   570..742
                     /product="growth hormone receptor"
     exon   570..742
                     /number=1
BASE COUNT   176 a   168 c   175 g   213 t
ORIGIN

Qy   1 CCTCCCCAAATCAAATTACATTTTCTC 26

Query Match   100.0%; Score 26; DB 4; Length 742;
Best Local Similarity   100.0%; Pred. No. 2.3;
Matches   26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CCTCCCCAAATCAAATTACATTTTCTC 26
       ||||||||||||||||||||||||||
Db 526 CCTCCCCAAATCAAATTACATTTTCTC 501

RESULT 2
LOCUS   HGU15741
DEFINITION Bos taurus somatotropin receptor gene, exon 1 and liver-specific
            promoter region.
ACCESSION   U15741
VERSION   U15741.2 GI:14570041
KEYWORDS   .
SOURCE   Bos taurus
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1 (bases 1 to 25688)
AUTHORS   Heap,D., Lucy,M.C., Collier,R.J., Boyd,C.K. and Warren,W.C.
TITLE   Rapid communication: nucleotide sequence of the promoter and first
            exon of the somatotropin receptor gene in cattle
JOURNAL   J. Anim. Sci. 74 (5), 1529 (1995)
MEDLINE   7665485
PUBMED   7665485
REFERENCE   2 (bases 1 to 25688)
AUTHORS   Lucy,M.C.
TITLE   Direct Submission
JOURNAL   Submitted (11-OCT-1994) Matthew C. Lucy, Animal Science, University
            of Missouri, 164 Animal Science Research Center, Columbia, MO
            65211, USA
REFERENCE   3 (bases 1 to 25688)
AUTHORS   Jiang,H. and Lucy,M.C.
TITLE   Involvement of hepatocyte nuclear factor 4 in the expression of the
            growth hormone receptor 1A messenger ribonucleic acid in bovine
            liver
JOURNAL   Mol. Endocrinol. 15 (6), 1023-1034 (2001)
MEDLINE   21270440
PUBMED   11376119
COMMENT   On Apr 10, 2001 this sequence version replaced gi:560024.
FEATURES   Location/Qualifiers
     source   1..25688
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="13-1"
                     /sex="male"

promoter   9763..10256
                     /tissue_type="liver"
                     /note="liver specific"
     repeat_region   10136..10175
                     /rpt_type=tandem
     TATA_signal   10232..10237
     mRNA   10257..>25688
                     /product="growth hormone receptor"
     exon   10258..10465
                     /note="liver-specific"
                     /number=1A
     exon   10192
                     /note="growth hormone receptor promoter; somatotropin receptor"
     CDS   25670..>25688
                     /codon_start=1
                     /product="growth hormone receptor"
                     /protein_id="AAK12298.1"
                     /db_xref="GI:13572470"
                     /translation="MDLNSC"
     exon   25660..>25688
                     /number=2
BASE COUNT   6552 a   5982 c   6605 g   7149 t
ORIGIN

Qy   1 CCTCCCCAAATCAAATTACATTTTCTC 26

Query Match   100.0%; Score 26; DB 4; Length 25688;
Best Local Similarity   100.0%; Pred. No. 1.5;
Matches   26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CCTCCCCAAATCAAATTACATTTTCTC 26
       ||||||||||||||||||||||||||
Db 10219 CCTCCCCAAATCAAATTACATTTTCTC 10194

RESULT 3
LOCUS   AF126288
DEFINITION Bos taurus growth hormone receptor (GHR) gene, promoter and 5'
            untranslated region.
ACCESSION   AF126288
VERSION   AF126288.1 GI:4894683
KEYWORDS   .
SOURCE   Bos taurus
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1 (bases 1 to 349)
AUTHORS   Ge,W., Davis,M.E., Hines,H.C. and Irvin,K.M.
TITLE   Two-allelic DGGE polymorphism detected in the promoter region of
            the bovine GHR gene
JOURNAL   Anim. Genet. 30 (1), 71 (1999)
MEDLINE   99159203
PUBMED   10050296
REFERENCE   2 (bases 1 to 349)
AUTHORS   Ge,W., Davis,M.E., Hines,H.C. and Irvin,K.M.
TITLE   Direct Submission
JOURNAL   Submitted (06-FEB-1999) Animal Sciences, The Ohio State University,
            2027 Coffey Road, Columbus, OH 43210, USA
FEATURES   Location/Qualifiers
     source   1..349
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /chromosome="20"
     gene   <1..>349
                     /gene="GHR"
     promoter   <1..302
                     /gene="GHR"
     variation   149
                     /gene="GHR"
                     /replace="a"
     mRNA   303..>349
                     /gene="GHR"

exon
   /gene="GHR"
   /product="growth hormone receptor"
   303. .>349
   /gene="GHR"
   /number=1
   303. .>349

5'UTR
   303. .>349
   /gene="GHR"

BASE COUNT   77 a   72 c   106 g   94 t
ORIGIN

Qy   1 CCTCCCCAAAT'AATTA'ATTTCTC 26
        ||| |||||||||||||||||||||
Db   259 CCTCGGCAAATCAATTACATTTCTC 234

Query Match
Best Local Similarity   87.7%;   Score 22.8;   DB 4;   Length 349;
Matches 24   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

RESULT 4
LOCUS   AC125465/c
DEFINITION   AC125465   14568 bp   DNA   linear   HTG 26-JUN-2002
SEQUENCE, 9 unordered pieces.
DEFINITION   Mus musculus chromosome UNK clone RP23-333A16, WORKING DRAFT
ACCESSION   AC125465
VERSION   AC125465.1 GI:2159226
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 14568)
AUTHORS   McPherson,J.D. and Waterston,R.H.
TITLE   The sequence of Mus musculus clone
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 14568)
AUTHORS   McPherson,J.D. and Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (26-JUN-2002) Genome Sequencing Center, 4444 Forest Park
             Parkway, St. Louis, MO 63108, USA

COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu

Center project name: M_BA0333A16

Summary Statistics

Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: phrap; version 0.990319
Consensus quality: 7558 bases at least Q40
Consensus quality: 9287 bases at least Q30
Consensus quality: 11208 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 13768; sum-of-contigs
Quality coverage: 0.09 in Q20 bases; agarose-fp
Quality coverage: 1.16 in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 9 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*   1   1015: contig of 1015 bp in length
*   1016 1115: gap of unknown length

FEATURES   Location/Qualifiers
     source   1. .14568
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-333A16"
     misc_feature   1. .1015
                     /note="assembly_name:Contig25"
     misc_feature   1116. .2365
                     /note="assembly_name:Contig33"
     misc_feature   2466. .3802
                     /note="assembly_name:Contig35"
     misc_feature   3903. .5693
                     /note="assembly_name:Contig38"
     misc_feature   5794. .7175
                     /note="assembly_name:Contig39"
     misc_feature   7276. .8375
                     /note="assembly_name:Contig40"
     misc_feature   8476. .9938
                     /note="assembly_name:Contig42"
     misc_feature   10039. .11427
                     /note="assembly_name:Contig43"
     misc_feature   11528. .14568
                     /note="assembly_name:Contig44"

*   1116: 2365: contig of 1250 bp in length
*   2366: 2465: gap of unknown length
*   2466: 3802: contig of 1337 bp in length
*   3802: 3902: gap of unknown length
*   3903: 5693: contig of 1791 bp in length
*   5693: 5793: gap of unknown length
*   5794: 7175: contig of 1382 bp in length
*   7175: 7275: gap of unknown length
*   7276: 8375: contig of 1100 bp in length
*   8375: 8475: gap of unknown length
*   8476: 9938: contig of 1463 bp in length
*   9938: 10038: gap of unknown length
*   10039: 11427: contig of 1389 bp in length
*   11427: 11527: gap of unknown length
*   11528: 14568: contig of 3041 bp in length

Qy   3 TCCCCAAATCAATTACATTTTCT 25
        |||||||||||||||||||| |||
Db 10186 TCCCCAAATCAATTACATATTCT 10164

Query Match
Best Local Similarity   82.3%;   Score 21.4;   DB 2;   Length 14568;
Matches 22   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

BASE COUNT   4173 a   2750 c   2822 g   4022 t   801 others
ORIGIN

RESULT 5
LOCUS   AC020937/c
ACCESSION   AC020937   94161 bp   DNA   linear   PRI 21-SEP-2001
DEFINITION   Homo sapiens chromosome 5 clone CTD-2239I2?, complete sequence.
ACCESSION   AC020937
VERSION   AC020937.5 GI:15718535
KEYWORDS   HTG.
SOURCE   Homo sapiens.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 94161)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 94161)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 94161)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE   Direct Submission
JOURNAL   Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 94161)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (21-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Sep 21, 2001 this sequence version replaced gi:15375144.
            Draft Sequence produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES   Location/Qualifiers
     source   1..94161
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
BASE COUNT   28997 a   19265 c   18272 g   27627 t
ORIGIN

Qy   4 CCCCAAATCAATTACATTTCTC 26
               ||||||||||||||||| |||||||
Db   77874 CCCCAAATCAATTAGATTTTTC 77852

Qy   4 CCCCAAATCAATTACATTTCTC 26
               ||||||||||||||||| |||||||
Db   8179 CCCCAAATCAATTAGATTTTTC 8157

Query Match   82.3%; Score 21.4; DB 9; Length 124495;
Best Local Similarity 95.7%; Pred. No. 87;
Matches   22; Conservative 0; Mismatches 1; Indels 0; Caps 0;

RESULT 6
ACCESSION   AC008502
LOCUS   AC008502   124495 bp   DNA   linear   PRI 20-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTC 1125:1, complete sequence.
VERSION   AC008502 R CI:14971159
KEYWORDS   HTG.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 124495)
AUTHORS   DOE Joint Genome Institute
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 124495)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (04-JUN-2001) DOE Joint Genome Institute and Stanford
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 124495)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (20-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Jul 20, 2001 this sequence replaced gi:14277251.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.4.
            STS Content:

SHGC-33971 G29509.   Location/Qualifiers
FEATURES   source   1..124495.
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
BASE COUNT   37253 a   25922 c   25561 g   35659 t
ORIGIN

Qy   4 CCCCAAATCAATTACATTTCTC 26
               ||||||||||||||||| |||||||
Db   8179 CCCCAAATCAATTAGATTTTTC 8157

Query Match   82.3%; Score 21.4; DB 9; Length 124495;
Best Local Similarity 95.7%; Pred. No. 84;
Matches   22; Conservative 0; Mismatches 1; Indels 0; Caps 0;

RESULT 7
ACCESSION   AC008787/c
LOCUS   AC008787   164478 bp   DNA   linear   HTG 05-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone CTD 2048E6, WORKING DRAFT SEQUENCE,
            29 unordered pieces.
VERSION   AC008787
ACCESSION   AC008787
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 164478)
AUTHORS   DOE Joint Genome Institute.
TITLE   Sequencing of Human Chromosome 5
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 164478)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On May 5, 2000 this sequence version replaced gi:5664425.
            --------------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 64224 t
            Center clone name: CITB-H1-2036E6

            Summary Statistics
            Consensus quality: 129975 bases at least Q40
            Consensus quality: 148252 bases at least Q30
            Consensus quality: 154251 bases at least Q20
            Quality coverage: 5.12 in Q20 bases; sum of contigs estimation
            Quality coverage: 4.95 in Q20 bases; sum of contigs estimation
            Estimated insert size: 162578; pulse field gel estimation
            Estimated insert size: 1.97000; pulse field gel estimation
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            ----
                      1   2262: contig of 2262 bp in length
                   2263   2362: gap of unknown length
                   2363   5469: contig of 3107 bp in length
                   5470   5569: gap of unknown length
                   5570   8642: contig of 3073 bp in length
                   8643   8742: gap of unknown length
                   8743   12920: contig of 4178 bp in length

```
 *    12921   13020: gap of unknown length
 *    13021   13620: contig of 3600 bp in length
 *    16621   16720: gap of unknown length
 *    16721   18847: contig of 2127 bp in length
 *    18848   18947: gap of unknown length
 *    18948   22210: contig of 3263 bp in length
 *    22211   22310: gap of unknown length
 *    22311   27037: contig of 4727 bp in length
 *    27038   27137: gap of unknown length
 *    27138   31662: contig of 4525 bp in length
 *    31663   31762: gap of unknown length
 *    31763   37223: contig of 5461 bp in length
 *    37224   37323: gap of unknown length
 *    37324   42764: contig of 5441 bp in length
 *    42765   42864: gap of unknown length
 *    42865   48849: contig of 5985 bp in length
 *    48850   48949: gap of unknown length
 *    48950   54895: contig of 5946 bp in length
 *    54896   54995: gap of unknown length
 *    54996   63708: contig of 8713 bp in length
 *    63709   63808: gap of unknown length
 *    63809   74072: contig of 10264 bp in length
 *    74073   74172: gap of unknown length
 *    74173   87006: contig of 12834 bp in length
 *    87007   87106: gap of unknown length
 *    87107   101078: contig of 13973 bp in length
 *    101079  101178: gap of unknown length
 *    101179  101780: contig of 12602 bp in length
 *    101781  101880: gap of unknown length
 *    113381  113880: contig of unknown length
 *    113782  113881: gap of unknown length
 *    113882  129485: contig of 15604 bp in length
 *    129486  129585: gap of unknown length
 *    129586  164478: contig of 34893 bp in length.
```

FEATURES   Location/Qualifiers
     source   1..164478
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2038E6"
                     /clone_lib="CalTech human BAC library D"

BASE COUNT   48924 a   32821 c   32766 g   48036 t   1937 others
ORIGIN

Qy   4 CCGCAAATCAATTACATTCTC 26
               |||||||||||| ||||||||
Db   84470 CCCTAAATCAATTACATTTTTC 84448

Query Match   82.3%;   Score 21.4;   DB 2;   Length 164478;
Best Local Similarity   95.7%;   Pred. No. 80;
Matches 22;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

RESULT 8
LOCUS   AC124754   186217 bp   DNA   linear   HTG 15 JUN 2002
DEFINITION Mus musculus chromosome UNK clone RP23-285C16, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC124754
VERSION   AC124754.1 GI:21428020
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 186217)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE   The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 186217)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT

                           ------- Genome Center -------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
                     -------- Project Information --------
            Center project name: M_BA0285C16
                     ------- Summary Statistics -------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye primer EP; 0% of reads
            Chemistry: Dye terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180867 bases at least Q40
            Consensus quality: 182198 bases at least Q30
            Consensus quality: 183076 bases at least Q20
            Insert size: 175000; agarose-fp
            Insert size: 185844; sum-of-contigs
            Quality coverage: 7.80 in Q20 bases; agarose-fp
            Quality coverage: 6.51 in Q20 bases; sum of contigs
     *   NOTE: This is a 'working draft' sequence. It currently
     *   consists of 14 contigs. The true order of the pieces
     *   is not known and their order in this sequence record is
     *   arbitrary. Gaps between the contigs are represented as
     *   runs of N, but the exact sizes of the gaps are unknown.
     *   This record will be updated with the finished sequence
     *   as soon as it is available and the accession number will
     *   be preserved.
     *   1   1495: contig of 1495 bp in length
     *   1496   1595: gap of unknown length.
     *   1596   3916: contig of 2321 bp in length
     *   3917   4016: gap of unknown length.
     *   4017   7041: contig of 3025 bp in length
     *   7042   7141: gap of unknown length
     *   7142   9838: contig of 2697 bp in length
     *   9839   9938: gap of unknown length.
     *   9939   16233: contig of 6295 bp in length
     *   16234   16333: gap of unknown length
     *   16334   23315: contig of 6982 bp in length
     *   23316   23415: gap of unknown length
     *   23416   34801: contig of 11386 bp in length
     *   34802   34901: gap of unknown length.
     *   34902   51362: contig of 16461 bp in length
     *   51363   51462: gap of unknown length
     *   51463   73102: contig of 21640 bp in length
     *   73103   73202: gap of unknown length
     *   73203   98846: contig of 25644 bp in length
     *   98847   98946: gap of unknown length
     *   98947   117407: contig of 18461 bp in length
     *   117408   117507: gap of unknown length
     *   117508   154549: contig of 37042 bp in length
     *   154550   154649: gap of unknown length
     *   154650   185986: contig of 31337 bp in length
     *   185987   186086: gap of unknown length
     *   186087   186217: contig of 131 bp in length.

FEATURES   Location/Qualifiers
     source   1..186217
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-285C16"
     misc_feature   1..1495
                     /note="assembly_name:Contig45"
     misc_feature   1596..3916
                     /note="assembly_name:Contig46"
     misc_feature   4017..7041
                     /note="assembly_name:Contig47"
     misc_feature   7142..9838
                     /note="assembly_name:Contig48"
     misc_feature   9939..16233
                     /note="assembly_name:Contig49"

```
LOCUS       AC069044                                      187592 bp    DNA    linear   HTG 04-MAY-2001
DEFINITION  Homo sapiens clone RP11-456A17, WORKING DRAFT SEQUENCE, 29
            unordered pieces.
ACCESSION   AC069044
VERSION     AC069044.2  GI:8699772
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187592)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-456A17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187592)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Bastien,N., Basten,F.,
            Benamara,A., Bloom,T., Brown,D.A., Burke,G.,
            Chooyote,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,J., DeArellano,K.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,J.,
            Grand Pierre,N., Grant,G., Hagos,B., Healford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Liu,C., Liu,G., Locke,C., Macdonald,J., Marquis,N.,
            McCarthy,M., McEwan,P., McCure,K., McKernan,K., Meldrim,J.,
            Meldrim,J., Merkel,J., Milnova,M., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,D., Pollara,V., Raymond,C., Riley,R., Rogov,P., Roltman,S.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,H.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trujillo,J.,
            Vassiliev,H., Vierl,R., V.S., Wilson,R., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 24, 2000 this sequence version replaced gi:7884563.
            All repeats were identified using RepeatMasker;
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/cgi-bin/RepeatMasker.html
            --------------  Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            --------------  Project Information
               Center project name: 456_A_17
               Center clone name: 456_A_17
            --------------  Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 173381 bases at least Q40
               Consensus quality: 180464 bases at least Q30
               Consensus quality: 184190 bases at least Q20
               Insert size: 188000; agarose-fp
               Insert size: 184892; sum-of-contigs
               Quality coverage: 8.9 in Q20 bases; agarose-fp
               Quality coverage: 8.9 in Q20 bases; sum of contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *               1 -  1249: contig of 1249 bp in length
            *            1338:  1338: gap of 100 bp
            *            1339 -  2646: contig of 1308 bp in length
            *            2647:  2746: gap of 100 bp
            *            2747 -  3747: contig of 1001 bp in length
            *            3748:  3847: gap of 100 bp
            *            3848 -  5418: contig of 1571 bp in length
            *            5419:  5518: gap of 100 bp
            *            5519 -  7079: contig of 1561 bp in length
            *            7080:  7179: gap of 100 bp
            *            7180 -  9028: contig of 1849 bp in length
            *            9029:  9128: gap of 100 bp
            *            9129 - 11378: contig of 2250 bp in length
            *           11379: 11478: gap of 100 bp
            *           11479 - 13484: contig of 2006 bp in length
            *           13485: 13584: gap of 100 bp
            *           13585 - 15737: contig of 2153 bp in length
            *           15738: 15837: gap of 100 bp
            *           15838 - 19964: contig of 4127 bp in length
            *           19965: 20064: gap of 100 bp
            *           20065 - 24312: contig of 4248 bp in length
            *           24313: 24412: gap of 100 bp
            *           24413 - 30036: contig of 5624 bp in length
            *           30037: 30136: gap of 100 bp
            *           30137 - 35281: contig of 5145 bp in length
            *           35282: 35381: gap of 100 bp
            *           35382 - 40530: contig of 5149 bp in length
            *           40531: 40630: gap of 100 bp
            *           40631 - 47249: contig of 6619 bp in length
            *           47250: 47349: gap of 100 bp
            *           47350 - 53028: contig of 5679 bp in length
            *           53029: 53128: gap of 100 bp
            *           53129 - 60168: contig of 7040 bp in length
            *           60169: 60268: gap of 100 bp
            *           60269 - 66601: contig of 6333 bp in length
            *           66602: 66701: gap of 100 bp
            *           66702 - 73033: contig of 6332 bp in length
            *           73034: 73133: gap of 100 bp
            *           73134 - 79683: contig of 6550 bp in length
            *           79684: 79783: gap of 100 bp
            *           79784 - 87783: contig of 8000 bp in length
```

```
FEATURES             Location/Qualifiers
source               1..187592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-456A17"
                     /clone_lib="RPCI-11 Human Male BAC"
misc_feature         1..1238
                     /note="assembly_fragment"
misc_feature         1339..2746
                     /note="assembly_fragment"
misc_feature         2747..3747
                     /note="assembly_fragment"
                     /insert_end:T7
                     clone_end:T7
                     vector_side:lett
misc_feature         3848..5418
                     /note="assembly_fragment"
misc_feature         5519..7079
                     /note="assembly_fragment"
misc_feature         7180..9028
                     /note="assembly_fragment"
misc_feature         9129..11178
                     /note="assembly_fragment"
misc_feature         11479..13484
                     /note="assembly_fragment"
misc_feature         13585..15737
                     /note="assembly_fragment"
misc_feature         15838..19964
                     /note="assembly_fragment"
misc_feature         20065..24312
                     /note="assembly_fragment"
misc_feature         24413..30036
                     /note="assembly_fragment"
misc_feature         30137..35281
                     /note="assembly_fragment"
misc_feature         35382..40530
                     /note="assembly_fragment"
misc_feature         40631..47249
                     /note="assembly_fragment"
misc_feature         47350..53028
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:left
misc_feature         53129..60168
                     /note="assembly_fragment"
misc_feature         60269..66601
                     /note="assembly_fragment"
misc_feature         66702..73033
                     /note="assembly_fragment"
misc_feature         73134..79683
                     /note="assembly_fragment"
misc_feature         79784..8783
                     /note="assembly_fragment"
misc_feature         87884..9500
                     /note="assembly_fragment"
misc_feature         9501..106227
                     /note="assembly_fragment"
misc_feature         10632?..115811
                     /note="assembly_fragment"
misc_feature         115912..125733
                     /note="assembly_fragment"
misc_feature         125834..138211
                     /note="assembly_fragment"
                     138312..161514
                     /note="assembly_fragment"
misc_feature         161615..187592
                     /note="assembly_fragment"
```

```
*    87784..87883: gap of         100 bp
*    87884..95000: contig of 7117 bp in length
*    95001..95100: gap of         100 bp
9501..106227: contig of 11127 bp in length
106228..106327: gap of         100 bp
106328..115811: contig of 9484 bp in length
115812..115911: gap of         100 bp
115912..125733: contig of 9822 bp in length
125734..125833: gap of         100 bp
125834..138211: contig of 12378 bp in length
138212..138311: gap of         100 bp
138312..161514: contig of 23203 bp in length
161515..161614: gap of         100 bp
161615..187592: contig of 25978 bp in length.
```

```
ORIGIN

BASE COUNT   54419 a   38222 c   37857 g   54441 t

                 /note="assembly_fragment"
misc_feature     161615..187592
                 /note="assembly_fragment"
misc_feature     138312..161514
                 /note="assembly_fragment"
misc_feature     125834..138211
                 /note="assembly_fragment"
```

RESULT 10
AC123369
LOCUS   AC123369   186653 bp   DNA   linear   HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-1cfL4, *** SEQUENCING IN PROGRESS
           ***, 66 unordered pieces.
ACCESSION AC123369
VERSION   AC123369.2 GI:21903235
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Rattus norvegicus.
ORGANISM   Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 186653)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
           Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
           Harbaria,J., Benton,J., Himage,K., Blankenburg,K., Bonnin,D.,
           Houck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
           Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
           Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
           Chen,G., Chen,P., Chen,Z., Chowdhry,I., Christopoulos,C.,
           Cleveland,C.D., Cox,C.V., Coyle,M.D., Dathorne,S.R., David,R.,
           Davila,M.L., Davis,C., Davy-Carroll,L., Deford,D.A.,
           Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
           Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
           Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
           Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
           Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
           Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
           Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
           Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,H.,
           Homsi,F., Howard,S., Huber,T., Hulyk,S., Hume,J., Jackson,L.E.,
           Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Jordan,S.,
           Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
           Kratovic,J., Kureshi,A., Landry,N., Leal,H., Lewis,L.C., Lewis,L.,
           Li,J., Li,Z., Liewl-arge,Q., Liu,J., Liu,W., Louisseged,H.,
           Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
           Maheshwari,M., Mapua,P., Martin,R., Martinez,E., Martinez,M.,
           Massey,E., Mawhiney,B., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
           Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
           Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
           Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
           Oragunye,N., Oviedo,R., Pace,A., Parten,A., Peery,J., Petri,J.,
           Peters,L., Pickens,R., Primus,E., Puie,M., Quiles,M., Ren,Y.,
           Fives,M., Rojas,A., Rodri-guez,D., Rolfe,M., Ruiz,S., Savery,G.,
           Scherer,S., Scott,G, Shen,H, Shooshari N. Sisson,I.,
           Soderling,F., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
           Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Taylor,H., Tang,H.,
           Tansey,J., Taylor,C., Taylor,J., Teiford,E., Thomas,N., Thomas,S.,
           Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,P., Wang,Q.,
           Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watley,K.,
           Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
           Wu,C., Wu,Y., Wu,Y., Zhou,J., Zortilla,S., Nelson,D.,
           Weinstock,G. and Gibbs,R.

```
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 186654)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
REFERENCE  4  (bases 1 to 186654)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Jul 18, 2002 this sequence version replaced gi:21240271.
           Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu
           Contact: hgsc-help@bcm.tmc.edu
              Project Information
           Center project name: GWFN
           Center clone name: CH240-122F14
              Summary Statistics
           Sequencing vector: Plasmid;
           Chemistry: Dye-terminator Big Dye; 1998 of reads
           Assembly program: Phrap; version 0.990329
           Consensus quality: 129877 bases at least Q40
           Consensus quality: 137214 bases at least Q30
           Consensus quality: 141956 bases at least Q20
           NOTE: Estimated insert size may differ from sequence length
           (see http://www.hgsc.bcm.tmc.edu/documentation.html)
           NOTE: This is a 'working draft' sequence. It currently
           consists of 66 contigs. The true order of the pieces
           is not known and their order in this sequence record is
           arbitrary. Gaps between the contigs are represented as
           runs of N, but the exact sizes of the gaps are unknown.
           This record will be updated with the finished sequence
           as soon as it is available and the accession number will
           be preserved.
```

*	1	1095: contig of 1095 bp in length
*	1096	1195: gap of unknown length
*	1196	2207: contig of 1012 bp in length
*	2208	2397: gap of unknown length
*	2398	3492: contig of 1185 bp in length
*	3493	4592: gap of unknown length
*	4593	4680: contig of 1088 bp in length
*	4681	4780: gap of unknown length
*	4781	6298: contig of 1518 bp in length
*	6299	6398: gap of unknown length
*	6399	7792: contig of 1394 bp in length
*	7793	7892: gap of unknown length
*	7893	9857: contig of 1965 bp in length
*	9858	9957: gap of unknown length
*	9958	11166: contig of 1409 bp in length
*	11167	11466: gap of unknown length
*	11467	12793: contig of 1327 bp in length
*	12794	12893: gap of unknown length
*	12894	14385: contig of 1492 bp in length
*	14386	14485: gap of unknown length
*	14486	15552: contig of 1067 bp in length
*	15553	15652: gap of unknown length
*	15653	17523: contig of 1871 bp in length
*	17524	17623: gap of unknown length
*	17624	19242: contig of 1619 bp in length
*	19243	19442: gap of unknown length
*	19443	20706: contig of 1164 bp in length
*	20707	20806: gap of unknown length
*	20807	22586: contig of 1780 bp in length
*	22587	22686: gap of unknown length
*	22687	24952: contig of 2266 bp in length
*	24953	25052: gap of unknown length

*	25053	27091: contig of 2039 bp in length
*	27092	27191: gap of unknown length
*	27192	29146: contig of 1955 bp in length
*	29147	29246: gap of unknown length
*	29247	30531: contig of 1285 bp in length
*	30532	30631: gap of unknown length
*	30632	32232: contig of 1601 bp in length
*	32233	32332: gap of unknown length
*	32333	33819: contig of 1487 bp in length
*	33820	33919: gap of unknown length
*	33920	36104: contig of 2185 bp in length
*	36105	36204: gap of unknown length
*	36205	37499: contig of 1295 bp in length
*	37500	37599: gap of unknown length
*	37600	39486: contig of 1887 bp in length
*	39487	39586: gap of unknown length
*	39587	41679: contig of 2093 bp in length
*	41680	41779: gap of unknown length
*	41780	43567: contig of 1788 bp in length
*	43568	43667: gap of unknown length
*	43668	46253: contig of 2586 bp in length
*	46254	46353: gap of unknown length
*	46354	48581: contig of 2228 bp in length
*	48582	48681: gap of unknown length
*	48682	50072: contig of 1391 bp in length
*	50073	50172: gap of unknown length
*	50173	51550: contig of 1378 bp in length
*	51551	51650: gap of unknown length
*	51651	53292: contig of 1642 bp in length
*	53293	53392: gap of unknown length
*	53393	54985: contig of 1593 bp in length
*	54986	55085: gap of unknown length
*	55086	56986: contig of 1901 bp in length
*	56987	57086: gap of unknown length
*	57087	59234: contig of 2148 bp in length
*	59235	59334: gap of unknown length
*	59335	61169: contig of 1835 bp in length
*	61170	61269: gap of unknown length
*	61270	63869: contig of 2600 bp in length
*	63870	63969: gap of unknown length
*	63970	65838: contig of 1869 bp in length
*	65839	65938: gap of unknown length
*	65939	67740: contig of 1801 bp in length
*	67741	67831: gap of unknown length
*	67832	70205: contig of 2374 bp in length
*	70206	70305: gap of unknown length
*	70306	72625: contig of 2320 bp in length
*	72626	72725: gap of unknown length
*	72726	75701: contig of 2976 bp in length
*	75702	75801: gap of unknown length
*	75802	79446: contig of 2545 bp in length
*	79447	79446: gap of unknown length
*	79447	82529: contig of 3084 bp in length
*	82530	82630: gap of unknown length
*	82631	85382: contig of 2752 bp in length
*	85383	85482: gap of unknown length
*	85483	87659: contig of 2177 bp in length
*	87660	87759: gap of unknown length
*	87760	90866: contig of 3107 bp in length
*	90867	90966: gap of unknown length
*	90967	94513: contig of 3547 bp in length
*	94514	94513: gap of unknown length
*	94514	96947: contig of 2434 bp in length
*	96948	97047: gap of unknown length
*	97048	99687: contig of 2640 bp in length
*	99688	99787: gap of unknown length
*	99788	102905: contig of 3118 bp in length
*	102906	103005: gap of unknown length
*	103006	106489: contig of 3484 bp in length
*	106490	106589: gap of unknown length
*	106590	110011: contig of 3422 bp in length
*	110012	110111: gap of unknown length
*	110112	113307: contig of 3196 bp in length

```
Qy    1 CCTCCCCAAATCAAATTACATTTCTCC 26
        | | | ||||||||||||||||||||||
Db 17793 CTTTCCAAAATCAAATTACATTTCTCC 17818
```

```
*    11330A    11407: gap of unknown length
*    11340B    117484: contig of 4077 bp in length

Query Match    88.58:  Score 21.2; DB 2; length 86653,
Best Local Similarity 88.58;  Pred. No. 95,
Matches  23; Conservative  0; Mismatches  3; Indels  0; Gaps  0;
```

```
RESULT 11
AC110785
LOCUS        AC110785          100987 bp   DNA   linear   PRI 09-MAR-2002
DEFINITION   Homo sapiens BAC clone RP11-45517 from 4, complete sequence.
ACCESSION    AC110785 AC044831
VERSION      AC110785.2 GI:18874242
KEYWORDS     HTG.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 100987)
AUTHORS      Wang,C., Haglund,K. and Kozlowicz,A.
TITLE        The sequence of Homo sapiens BAC clone RP11-45517
JOURNAL      Unpublished (2001)
REFERENCE    2  (bases 1 to 100987)
AUTHORS      Sulston,J.E and Waterston,R
TITLE        Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99069792
PUBMED       9847074
REFERENCE    3  (bases 1 to 100987)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (15-FEB-2002) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE    4  (bases 1 to 100987)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (16-FEB-2002) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St Louis,
             MO 63108, USA
REFERENCE    5  (bases 1 to 100987)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (24 FEB 2002) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE    6  (bases 1 to 100987)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (09-MAR-2002) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      On Feb 24, 2002 this sequence version replaced gi:18677616.
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc
             Contact: sapiens@watson.wustl.edu
             --------------
             Summary Statistics
             Center project name: H_NH0455107
             Drafting Center: WIBR

             NOTICE: This sequence may not represent the entire insert of this
             clone. It may be shorter because we only sequence overlapping
             clone sections once, or longer because we provide a small overlap
             between neighboring data submissions.

             This sequence was finished as follows unless otherwise noted:
             all regions were double stranded, sequenced with an alternate
             chemistry, or covered by high quality data (i.e., phred quality >-
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by sequence
             from more than one subclone; and the assembly was confirmed by
             restriction digest.

             MAPPING INFORMATION:
             Mapping information for this clone was provided by Dr. John D.
             McPherson, Department of Genetics, Washington University, St. Louis
             MO. For additional information about the map position of this
             sequence, see http://genome.wustl.edu/gsc

             SOURCE INFORMATION:
             The RP11-11 human BAC library was made from the blood of one male
             donor, as described by Osoegawa,K., Woon,P.Y., Zhao,H., Frengen,E.,
             Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
             approach for construction of bacterial artificial chromosome
             libraries. Genomics 51:1-8. The clone may be obtained either from
             Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
             and coworkers at http://www.chori.org
             VECTOR: pBACe3.6

             NEIGHBORING SEQUENCE INFORMATION:
             The clone sequenced to the left is RP11-507B10, 2000 bp overlap;
             the clone sequenced to the right is RP11 435C3, 2000 bp overlap.
             Actual start of this clone is at base position 34465 of
             RP11-507B10.

             Polymorphisms exists between AC110785 and AC020556.

             The sequence of AC044831 has been incorporated into AC110785.

FEATURES             Location/Qualifiers
     source          1..100987
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-45517"
                     /clone_lib="RPCI-11"
     repeat_region   126..873
                     /rpt_family="ERVL"
     repeat_region   1260..1831
                     /rpt_family="ERVL"
     repeat_region   4910..5211
                     /rpt_family="Alu"
     repeat_region   5274..5302
                     /rpt_family="AT_rich"
     repeat_region   5319..5621
                     /rpt_family="Alu"
     repeat_region   5634..5663
                     /rpt_family="Alu"
     repeat_region   5686..5758
                     /rpt_family="AT_rich"
     repeat_region   5997..6368
                     /rpt_family="(TA)n"
     repeat_region   6407..6440
                     /rpt_family="MaLR"
     repeat_region   6575..6598
                     /rpt_family="AT_rich"
     repeat_region   7028..7053
                     /rpt_family="AT_rich"
     repeat_region   7054..7380
                     /rpt_family="Alu"
     repeat_region   9177..9632
                     /rpt_family="Alu"
     repeat_region   9937..10192
                     /rpt_family="ERVL"
     repeat_region   10225..10888
                     /rpt_family="L1"
     repeat_region   11315..11433
                     /rpt_family="L1,2"
     repeat_region   11434..11512
                     /rpt_family="L1"
```

```
repeat_region    11810..11944    /rpt_family "Alu"
repeat_region    13466..13950    /rpt_family "L1"
repeat_region    14068..14144    /rpt_family "L1"
repeat_region    14183..14479    /rpt_family "AT_rich"
repeat_region    14548..14664    /rpt_family "L1"
repeat_region    14699..15082    /rpt_family "L1"
repeat_region    15106..16815    /rpt_family "L1"
repeat_region    18895..20357    /rpt_family "L1"
repeat_region    20458..20603    /rpt_family "Alu"
repeat_region    20604..21627    /rpt_family "Alu"
repeat_region    21628..21924    /rpt_family "L1"
repeat_region    21809..24139    /rpt_family "Alu"
repeat_region    23925..22649    /rpt_family "L1"
repeat_region    24574..24885    /rpt_family "Alu"
repeat_region    22781..22812    /rpt_family "Alu"
repeat_region    24923..25082    /rpt_family "L1"
repeat_region    23216..23511    /rpt_family "AT_rich"
repeat_region    25549..25646    /rpt_family "L1"
repeat_region    25653..25959    /rpt_family "MER1_type"
repeat_region    2903..29448     /rpt_family "Alu"
repeat_region    31016..32059    /rpt_family "L1"
repeat_region    12164..32210    /rpt_family "L1"
repeat_region    32519..32876    /rpt_family "L2"
repeat_region    40010..40109    /rpt_family "MaLR"
repeat_region    40124..40224    /rpt_family "L2"
repeat_region    28877..28877    /rpt_family "MaLR"
repeat_region    13850..34293    /rpt_family "L2"
repeat_region    34297..34621    /rpt_family "MaLR"
repeat_region    44968..45165    /rpt_family "MaLR"
repeat_region    45196..45280    /rpt_family "MIR"
repeat_region    46354..46374    /rpt_family "L2"
repeat_region    46395..46601    /rpt_family "AU_rich"
repeat_region    46695..46609    /rpt_family "L1"
repeat_region                    47585
```

Query Match   77.7%;   Score 20.2;   DB 9;   Length 100987;
Host Local Similarity   88.0%;   Pred. No. 2.5e-02;
Matches   22;   Conservative   0;   Mismatches   3;   Indels   0;   Caps   0;

```
Qy      1   CCTCCCCAAATCAATTACATTTCT  25
            ||||||||||||||||||||||||
Db  74441   CATCCCCAAATCAATTCCATTTGCT  74465
```

RESULT 12
Ap004970

LOCUS   AP004970   106000 bp   DNA   linear   PLN 19 JUL 2002
DEFINITION   Lotus japonicus genomic DNA, chromosome 5, clone:LjT30P03, TM0148,
              complete sequence.
ACCESSION   AP004970
VERSION   AP004970.1   GI:21907988
KEYWORDS   HTG.
SOURCE   Lotus japonicus
  ORGANISM   Lotus japonicus
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
              Lotus.
REFERENCE   1
  AUTHORS   Kato,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S., and Tabata,S.
  TITLE   Structural Analysis of a Lotus japonicus Genome. I. Sequence
              Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb
              Regions of the Genome
  JOURNAL   Unpublished
REFERENCE   2   (bases 1 to 106000)
  AUTHORS   Nakamura,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research
              Institute, Department of Plant Gene Research; 1532-3, Yana,
              Kisarazu, Chiba 292-0812, Japan (E-mail:; kazusa.or.jp, Tel:81-438-52-3935,
              URL:http://www.kazusa.or.jp; Tel:81-438-52-3935,
              Fax:81-438-52-3934)
FEATURES   Location/Qualifiers
     source   1..106000
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="5"
                     /clone="LjT40P03"
                     /clone="LjT40P03"
                     /clone="LjT library"
                     /note="TAC clone:TM0148"
BASE COUNT   33621 a   19351 c   18804 g   14224 t
ORIGIN
```

```
Qy      1   CCTCCGCAAATCAATACCCTTTCT  25
            |||| ||||||||||  ||||||||
Db  77429   CCTCCCCAAATCAATTGCCTTTCT  77453
```

Query Match   77.7%;   Score 20.2;   DB 8;   Length 106000;
Host Local Similarity   88.0%;   Pred. No. 2.6e-02;
Matches   22;   Conservative   0;   Mismatches   3;   Indels   0;   Caps   0;

RESULT 13
Ac073450/6

LOCUS   AC073450   167623 bp   DNA   linear   HTG 04-OCT 2000
DEFINITION   Homo sapiens chromosome 4 clone RP11-521C6 map 4, WORKING DRAFT
              SEQUENCE, 21 unordered pieces.
ACCESSION   AC073450
VERSION   AC073450.2   GI:10567951
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1   (bases 1 to 167623)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., and Lander,E.
  TITLE   Homo sapiens chromosome 4, clone RP11-521C6
  JOURNAL   Unpublished
REFERENCE   2   (bases 1 to 167623)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H.,
              Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
              Boguslavliy,L., Boukbgalter,B., Brown,A., Burkett,G.,
```

```
            Campopiano,A., Castle,A., Chuprel,V., Cuiangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,L.S.,
         *  Dodge,S., Domino,M., Doyle,M., Ferreira,J., Fitzhugh,W., Gage,D.,
         *  Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
         *  Grand-Pierre,M., Grant,G., Hagos,B., Heaford,A., Horton,L.,
         *  Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kahn,L., Karatas,A.,
         *  Kcir,F., Lakonare,K., Landers,T., Landers,T., Leczky,J.,
         *  Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,F., Marquis,N.,
         *  McCarthy,M., McEwan,P., McKain,A., McKhann,K., McPheeters,R.,
         *  Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
         *  Murphy,T., Najmi,C.H., Nguyen,C., Nicol,R., Oberholf,F.,
         *  O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
         *  Pisani,C., Pollara,V., Raymond,C., Riley,P., Rogov,P., Rothman,D.,
         *  Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
         *  Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
         *  Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trijillo,T.,
         *  Vassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
         *  Young,G., Zainoun,J., Zimmer,A. and Zody,M.
         *  ------
         *           Genome Center
            COMMENT     Center: Whitehead Institute/ MIT Center for Genome Research
                        Center code: WIBR
                        Web site: http://www-seq.wi.mit.edu
                        Contact: sequence_submissions@genome.wi.mit.edu
                        ------
                           Project Information
                        Center project name: L10444
                        Center clone name: 521_C_6
                        ------
                           Summary Statistics
                        Sequencing vector: M13; M77815; 100% of reads
                        Chemistry: Dye terminator Big Dye; 100% of reads
                        Assembly program: Phrap; version 0.960731
                        Consensus quality: 15720 bases at least Q40
                        Consensus quality: 162563 bases at least Q30
                        Consensus quality: 164497 bases at least Q20
                        Insert size: 165623; sum-of-contigs
                        Quality coverage: 4.3 in Q20 bases; sum-of-contigs
            ------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            ------
            *      1     586: contig of 586 bp in length
            *    587     686: gap of           100 bp
            *    687    1860: contig of 1174 bp in length
            *   1861    1960: gap of           100 bp
            *   1961    2359: contig of 1599 bp in length
            *   3560    3659: gap of           100 bp
            *   3560   31851: contig of 28192 bp in length
            *  31852   31951: gap of           100 bp
            *  31952   33013: contig of 1062 bp in length
            *  33014   33113: gap of           100 bp
            *  33114   35199: contig of 2086 bp in length
            *  35200   35299: gap of           100 bp
            *  35300   38384: contig of 3085 bp in length
            *  38385   38484: gap of           100 bp
            *  38485   43409: contig of 4925 bp in length
            *  43410   43509: gap of           100 bp
            *  43510   48273: contig of 4764 bp in length
            *  48274   48373: gap of           100 bp
            *  48374   53158: contig of 4785 bp in length
            *  53159   53258: gap of           100 bp
            *  53259   60666: contig of 7408 bp in length
            *  60667   60766: gap of           100 bp
```

```
FEATURES             Location/Qualifiers
     source          1..167623
                     /organism="Homo sapiens"
                     /clone_end:left
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-521C6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..586
                     /note="assembly_fragment"
     misc_feature    687..1860
                     /note="assembly_fragment"
     misc_feature    1961..3559
                     /note="assembly_fragment"
     misc_feature    3660..31851
                     /note="assembly_fragment"
     misc_feature    31952..33013
                     /note="assembly_fragment"
     misc_feature    33114..35199
                     /note="assembly_fragment"
     misc_feature    35300..38384
                     /note="assembly_fragment"
     misc_feature    38485..43409
                     /note="assembly_fragment"
     misc_feature    43510..48273
                     /note="assembly_fragment"
     misc_feature    48374..53158
                     /note="assembly_fragment"
     misc_feature    53259..60666
                     /note="assembly_fragment"
     misc_feature    60767..68003
                     /note="assembly_fragment"
     misc_feature    68104..77478
                     /note="assembly_fragment"
     misc_feature    77579..86611
                     /note="assembly_fragment"
     misc_feature    86712..95048
                     /note="assembly_fragment"
     misc_feature    95149..107974
                     /note="assembly_fragment"
     misc_feature    108075..122534
                     /note="assembly_fragment"
     misc_feature    122635..137590
                     /note="assembly_fragment"
     misc_feature    137691..152722
                     /note="assembly_fragment"
     misc_feature    152823..167352
                     /note="assembly_fragment"
     misc_feature    167453..167623
                     /note="assembly_fragment"
                     /clone_end:T7
                     /vector_side:right
```

BASE COUNT
ORIGIN

Query Match   %2422 a 41487 c 41524 g 50287 t   2003 others
Best Local Similarity   77.7%;
Host   88.0%; Pred. N: 2.1e-02;
Matches 22; Conserved 0; Mismatches 4; Indels 0; Gaps 0;

Qy   2 GTCCCCAAATCCAATATAATTCCCCC 26
             ||||| ||||||||||||||||||
Db 167492 GTCAACAAATCCAATATAATTCCCCC 167468

RESULT 14
AC107H04/c
AC107H04/c
LOCUS   AC107H04   230419 bp   DNA   linear   HTG 21-AUG-2002
DEFINITION   Mus musculus clone RP23-48C16, WORKING DRAFT SEQUENCE, 6 ordered
             pieces.
ACCESSION   AC107H04
VERSION   AC107H04.4 GI:22981958
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 230419)
AUTHORS   Birren,B., Nusbaum,C., and Lander,E.
TITLE   Mus musculus, clone RP23-48C16
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 230419)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
             Anderson,S., Baldwin,J., Barnes,T., Bastien,V., Boguslavsky,J., Brokhkuiser,R.,
             Brown,A., Cahill,P., Campopiano,A., Chang,J., Chazaro,9.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
             Cooke,P., DeArellano,K., DeWar,K., Piaz,J., Podge,S., Fare,S.,
             Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Cardyna,S.,
             Gloede,S., Gurd,S., Goyette,M., Graham,L., Grand Pierre,N.,
             Hafez,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
             Kamat,A., Karafas,A., Kells,C., LaRocque,R., Lamazares,R.,
             Landers,T., Lebazky,T., Levine,R., Li,B., Maclean,C.,
             Macdonald,P., Major,J., Marquis,N., Matthews,N., McCarthy,M.,
             McGuard,P., McKernan,K., Meldrim,J., Mesrob,J., Mihova,T.,
             Melnoa,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
             Norman,C.H., O'Donnell,P., O'Neill,P., Oliver,J.,
             Peterson,M., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
             Ketta,K., Richback,M., Kiley,K., Rise,C., Rogov,P., Roman,J.,
             Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
             Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
             Strauss,J., Subramanian,A., Talamas,J., TesLaye,S., Theodore,J.,
             Topham,K., Travers,M., Travis,J., Triglllo,J., Vassiliev,H.,
             Viel,R., Vo,A., Wilson,B., Wyman,D., Ye,W., Young,G.,
             Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT   On Aug 21, 2002 this sequence version replaced gi:21126296.
             TITLE   Direct Submission
             JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
                          Research, 320 Charles Street, Cambridge, MA 02141, USA
             CR Aug 21, 2002 this sequence version replaced gi:21126296.
             All repeats were identified using RepeatMasker;
             Smit, A.F.A. & Green, P. (1996,1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html

             Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIHR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
             --------------- Project Information ---------------
             Center project name: L20527
             Center clone name: 48_C_16
             --------------- Summary Statistics ---------------
             Sequencing Vector: Plasmid; n/a; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.960731
             Consensus quality: 229651 bases at least Q40
             Consensus quality: 229730 bases at least Q30
             Consensus quality: 229770 bases at least Q20
             Insert size: 225000; agarose-fp
             Insert size: 229419; sum-of-contigs
             Quality coverage: 11.8 in Q40 bases; agarose-fp
             Quality coverage: 11.8 in Q20 bases; agarose-fp
             Quality coverage: 11.6 in Q20 bases; sum of contigs

             NOTE: This is a 'working draft' sequence. It currently
             consists of 6 contigs. Gaps between the contigs
             are represented as runs of N. The order of the pieces
             is believed to be correct as given, however the order
             of the gaps between them are based on estimates that have
             provided by the submitter.
             This sequence will be replaced
             by the finished sequence as soon as it is available and
             the accession number will be preserved.
             1   230419: contig of 1429 bp in length
             1429: gap of   100 bp
             1430   1529: contig of   100 bp in length
             1530   22415: gap of   20786 bp in length
             22416 22415: contig of   100 bp in length
             22416 77795: gap of   55480 bp in length
             77796 77895: contig of   100 bp in length
             77896 125553: gap of   47658 bp in length
             125554 125653: contig of   100 bp in length
             125654 188794: gap of   63141 bp in length
             188795 188894: contig of   100 bp in length
             188895 230419: contig of   41525 bp in length

FEATURES   Location/Qualifiers
     source   1..230419
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /clone="RP23-48C16"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1..1429
                     /note="assembly_fragment"
                     /clone_end:SP6
                     /vector_side:left"
     misc_feature   1530..22415
                     /note "assembly_fragment"
                     /vector_side:left"
     misc_feature   22416..77795
                     /note "assembly_fragment"
     misc_feature   77896..125553
                     /note "assembly_fragment"
     misc_feature   125654..188794
                     /note="assembly_fragment"
     misc_feature   188895..230419
                     /note="assembly_fragment"
                     /clone_end:T7
                     /vector_side:right"

BASE COUNT   59296 a   44015 c 43325 g   74208 t   545 others
ORIGIN

Query Match   77.7%; Score 20.2; DB 2; Length 230419;

```
Best local Similarity  89.0%, Conservative  0, Mismatches  3, Indels  0, Gaps  0;
Matches 22,                     and No 2 hp 02,

QY    1 CTCCCCAAATCAATTACAATTTCT 25
        ||||||||||||||||| |||||| |||
Db 165061 CTACCCCAAAATGAATTAACTTTGCT 165037

RESULT 15
AC016175/c
LOCUS       AC016175              95445 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-20M11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016175
VERSION     AC016175.2  GI:9123967
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 95445)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Unpublished
  JOURNAL
REFERENCE   2  (bases 1 to 95445)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,R., Donglan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,C., Hagos,B., Heaford,A., Horton,L.,
            Howland,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lui,A., Lokyitsang,T., MacDonald,M.,
            McEwan,P., McKernan,K., McLaughlin,S., Meldrim,J.,
            Morrow,T., Naylor,J., Nguyen,C., Nicol,R., Nusbaum,C.,
            Peterson,K., Pollara,V., Riley,F., Roy,A., Santos,P., Savery,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-Nov-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13 2000 this sequence version replaced gi:6466619.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information ----------
                Center project name: L3991
                Center clone name: 20_M11

            NOTE: This record contains 91 individual
            sequencing reads that have not been assembled into
            contigs. Runs of N are used to separate the reads
            and the order in which they appear is completely
            arbitrary. Low pass sequence sampling is useful for
            identifying clones that may be gene-rich and allows
            overlap relationships among clones to be deduced.
            However, it should not be assumed that this clone
            will be sequenced to completion. In the event that
            the record is updated, the accession number will
            be preserved.

                 1     943: contig of 943 bp in length
               944    1043: gap of 100 bp
              1044    2014: contig of 971 bp in length
              2015    2114: gap of 100 bp
              2115    3045: contig of 931 bp in length
              3046    3145: gap of 100 bp
              3146    4137: contig of 992 bp in length
              4138    4237: gap of 100 bp
```

```
              4238    5204: contig of 957 bp in length
              5205    5304: gap of 100 bp
              5305    6255: contig of 951 bp in length
              6256    6355: gap of 100 bp
              6356    7315: contig of 960 bp in length
              7316    7415: gap of 100 bp
              7416    8363: contig of 948 bp in length
              8364    8463: gap of 100 bp
              8464    9416: contig of 953 bp in length
              9417    9516: gap of 100 bp
              9517   10473: contig of 957 bp in length
             10474   10573: gap of 100 bp
             10574   11524: contig of 951 bp in length
             11525   11624: gap of 100 bp
             11625   12565: contig of 941 bp in length
             12566   12665: gap of 100 bp
             12666   13609: contig of 944 bp in length
             13610   13709: gap of 100 bp
             13710   14666: contig of 957 bp in length
             14667   14766: gap of 100 bp
             14767   15743: contig of 977 bp in length
             15744   15843: gap of 100 bp
             15844   16791: contig of 948 bp in length
             16792   16891: gap of 100 bp
             16892   17843: contig of 952 bp in length
             17844   17943: gap of 100 bp
             17944   18920: contig of 977 bp in length
             18921   19020: gap of 100 bp
             19021   19953: contig of 933 bp in length
             19954   20053: gap of 100 bp
             20054   20979: contig of 926 bp in length
             20980   21079: gap of 100 bp
             21079   22015: contig of 937 bp in length
             22016   22115: gap of 100 bp
             22116   23065: contig of 949 bp in length
             23066   23165: gap of 100 bp
             23166   24107: contig of 942 bp in length
             24108   24207: gap of 100 bp
             24208   25126: contig of 919 bp in length
             25127   25226: gap of 100 bp
             25227   26186: contig of 960 bp in length
             26187   26287: gap of 100 bp
             26287   27240: contig of 954 bp in length
             27241   27340: gap of 100 bp
             27341   28309: contig of 969 bp in length
             28310   28409: gap of 100 bp
             28410   29364: contig of 955 bp in length
             29365   29464: gap of 100 bp
             29465   30435: contig of 971 bp in length
             30436   30535: gap of 100 bp
             30536   31476: contig of 941 bp in length
             31477   31576: gap of 100 bp
             31577   32532: contig of 956 bp in length
             32533   32632: gap of 100 bp
             32633   33591: contig of 959 bp in length
             33592   33691: gap of 100 bp
             33692   34636: contig of 945 bp in length
             34637   34736: gap of 100 bp
             34737   35689: contig of 953 bp in length
             35690   35789: gap of 100 bp
             35790   36751: contig of 962 bp in length
             36752   36851: gap of 100 bp
             36852   37789: contig of 938 bp in length
             37790   37889: gap of 100 bp
             37890   38859: contig of 970 bp in length
             38860   38959: gap of 100 bp
             38960   39864: contig of 905 bp in length
             39865   39964: gap of 100 bp
             39965   40924: contig of 960 bp in length
             40925   41024: gap of 100 bp
             41025   41990: contig of 966 bp in length
             41991   42090: gap of 100 bp
             42091   43002: contig of 912 bp in length
```

Search completed: January 14, 2003, 14:08:57
Job time : 425.333 secs

```
44003  44102: gap of    100 bp in length
44103  44051: contig of  949 bp in length
44052  44151: gap of    100 bp in length
44152  44089: contig of  918 bp in length
45090  45189: gap of    100 bp in length
45190  46135: contig of  946 bp in length
46136  46235: gap of    100 bp in length
46236  47168: contig of  934 bp in length
47169  47268: gap of    100 bp in length
47269  48208: contig of  940 bp in length
48209  48408: gap of    100 bp in length
48409  49276: contig of  968 bp in length
49277  49377: gap of    100 bp in length
49377  50333: contig of  958 bp in length
50345  50444: gap of    100 bp in length
50445  51316: contig of  882 bp in length
51417  51416: gap of    100 bp in length
51417  52392: contig of  976 bp in length
52393  52492: gap of    100 bp in length
52493  53398: contig of  906 bp in length
53399  53498: gap of    100 bp in length
53499  54459: contig of  961 bp in length
54460  54559: gap of    100 bp in length
54559  55516: contig of  957 bp in length
55517  55616: gap of    100 bp in length
55617  56575: contig of  959 bp in length
56576  56675: gap of    100 bp in length
56676  57638: contig of  963 bp in length
57749  57748: gap of    100 bp in length
57749  58706: contig of  968 bp in length
58707  58806: gap of    100 bp in length
58807  59747: contig of  931 bp in length
59748  59847: gap of    100 bp in length
59848  60776: contig of  949 bp in length
60777  60876: gap of    100 bp in length
60877  61862: contig of  986 bp in length
61863  61962: gap of    100 bp in length
61963  62954: contig of  992 bp in length
62955  63054: gap of    100 bp in length
63054  64051: contig of  897 bp in length
64051  64951: gap of    100 bp in length
64955  65004: contig of  952 bp in length
65004  65103: gap of    100 bp in length
65103  66057: contig of  954 bp in length
66057  66157: gap of    100 bp in length
66158  67112: contig of  955 bp in length
67113  67212: gap of    100 bp in length
67213  68176: contig of  964 bp in length
68177  68276: gap of    100 bp in length
68277  69207: contig of  931 bp in length
69208  69307: gap of    100 bp in length
69307  70204: contig of  896 bp in length
70204  70303: gap of    100 bp in length
70304  71245: contig of  942 bp in length
71246  71345: gap of    100 bp in length
71346  72289: contig of  954 bp in length
72290  72389: gap of    100 bp in length
72390  73363: contig of  974 bp in length
73364  73463: gap of    100 bp in length
73464  74456: contig of  994 bp in length
74457  74556: gap of    100 bp in length
74557  75492: contig of  946 bp in length
75493  75592: gap of    100 bp in length
75493  75592: contig of  933 bp in length
75593  76525: gap of    100 bp in length
```

```
Query Match        76.28;  Score 19.8;  DB 2;  Length 95445;
Best Local Similarity  91.4%;
Matches  23;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy       3 TCGCTAATCAATTAGCATTTTCT 25
           ||||||||||||||||||| || ||
Db   63484 TCGCCAAATAATTACTTTTCU 63462
```